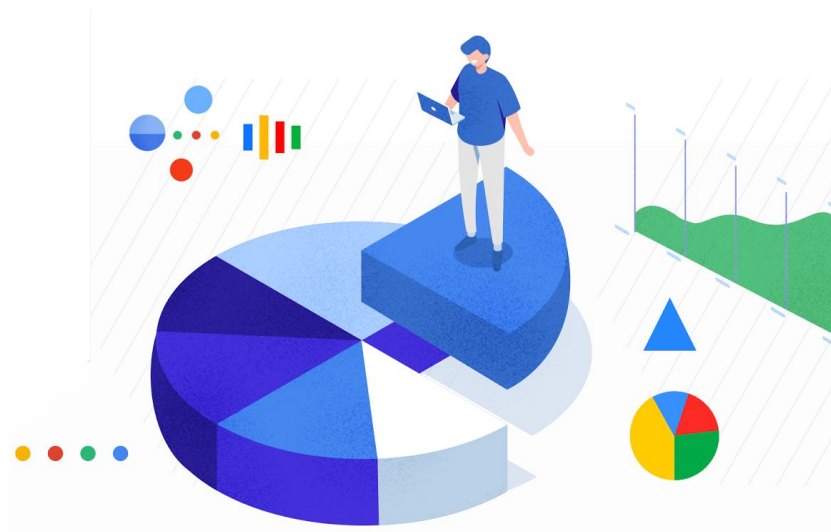


Complementing Data Ecosystems with Google Cloud

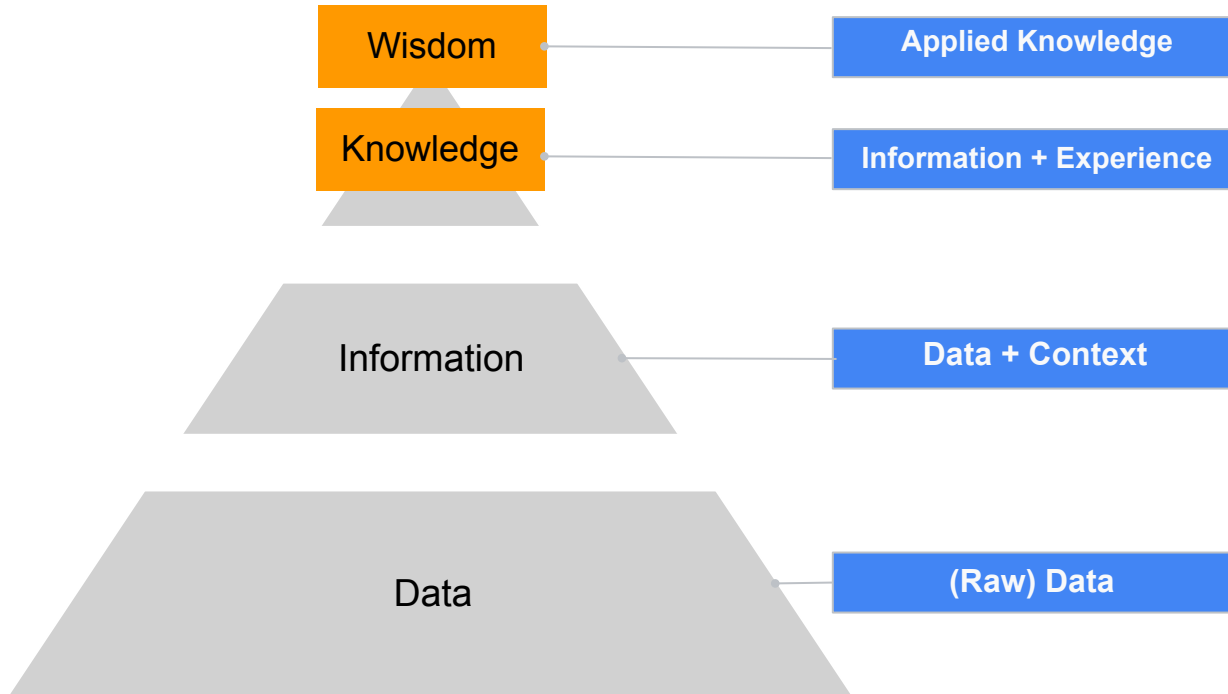


Data Spaces Symposium 2024, Darmstadt
Thomas Baumgart
Principal Architect, Google Cloud



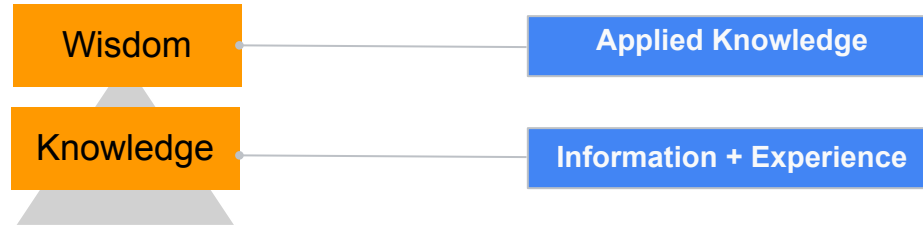
Information Management Processing

The Knowledge Pyramid - **Data alone has limited Value!**



Information Management Processing

The Knowledge Pyramid - **Data alone has no Value!**



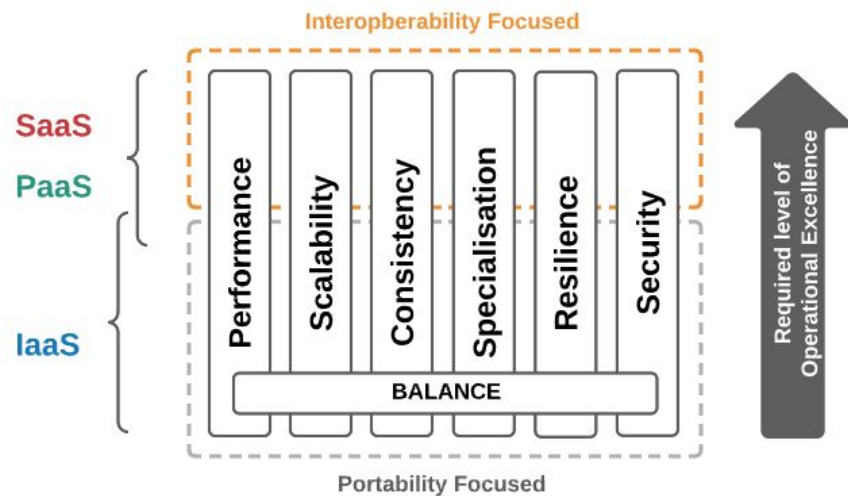
- **Advanced Data Processing** might in some cases **require more sophisticated PaaS and SaaS Services** -> introduction of **complementary “Interoperability focused” Services**
- **Data Space Core- / Platform- Services** require **highest level of Sovereignty** including **Software Sovereignty** -> **Portability is a major Architecture Principle**

Data + Context

(Raw) Data

Sophisticated AT SCALE services demand highly mature cloud operation models

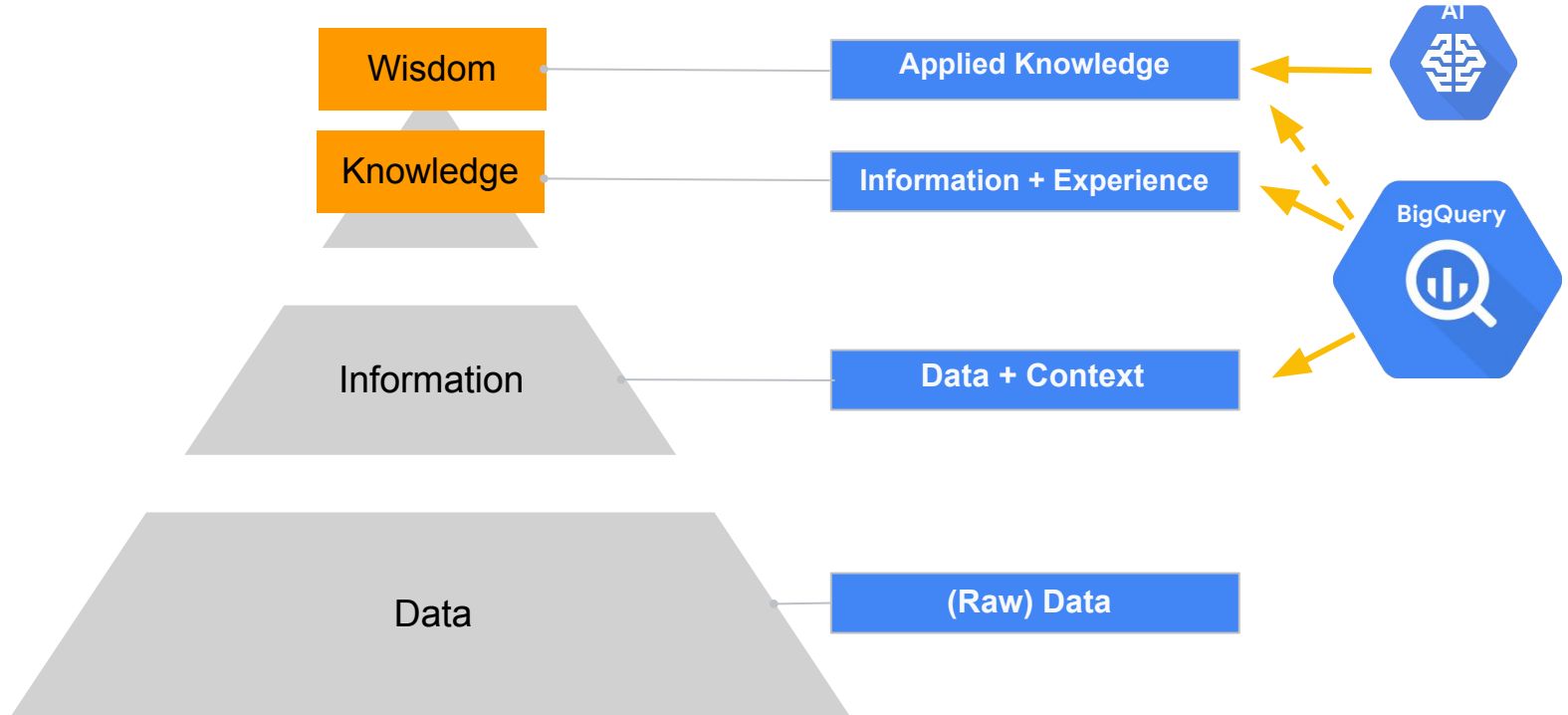
Containers are NOT the ultimate Answer to everything!



- The less **generic** (IaaS->SaaS) a **service** becomes, the less **portable it gets**
- **Operational Excellence** is the most **determining factor for highly productive Cloud Services**
- Highly **specialized and sophisticated PaaS or SaaS** services are **typically not portable** as they **require optimized infrastructure & operation** processes - so **focusing on Interoperability**

Information Management Processing

How to utilize State of the Art Cloud Services in Data Space Scenarios





BigQuery

Cloud Data Warehouse

Some BigQuery Stats

10.5 Trillion Largest query (rows)

5 petabytes Largest query (data size)

250 petabytes Largest storage customer

4.5 million rows/sec Peak ingestion rate

Google Cloud



Google Cloud's **enterprise data warehouse** for analytics

Gigabyte to **exabyte scale** storage and SQL queries

Encrypted, durable,
And highly available

Fully managed and **serverless** for maximum agility and scale

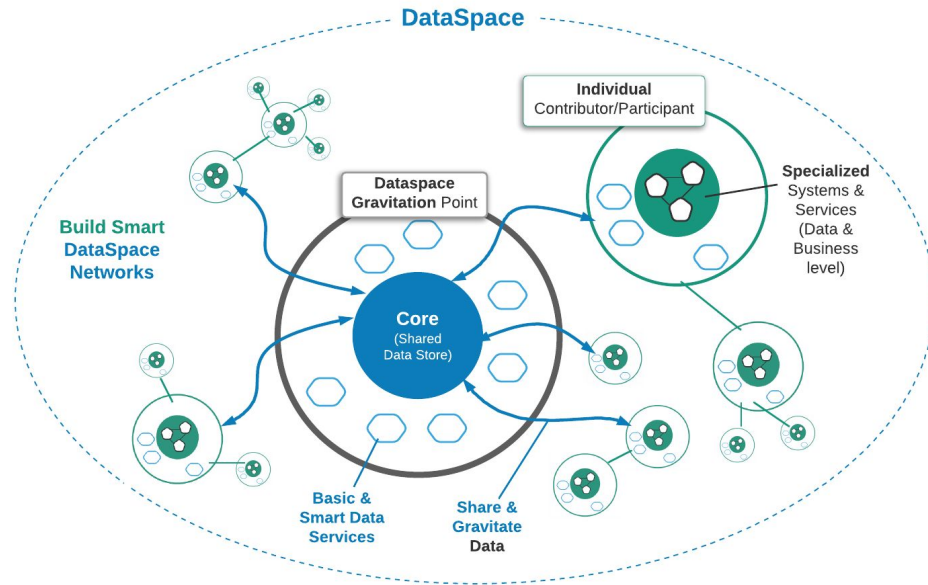
Real-time insights from streaming data

Built-in **ML** for out-of-the-box predictive insights

High-speed, in-memory **BI Engine** for faster reporting and analysis

Composing Smart Data Space Networks

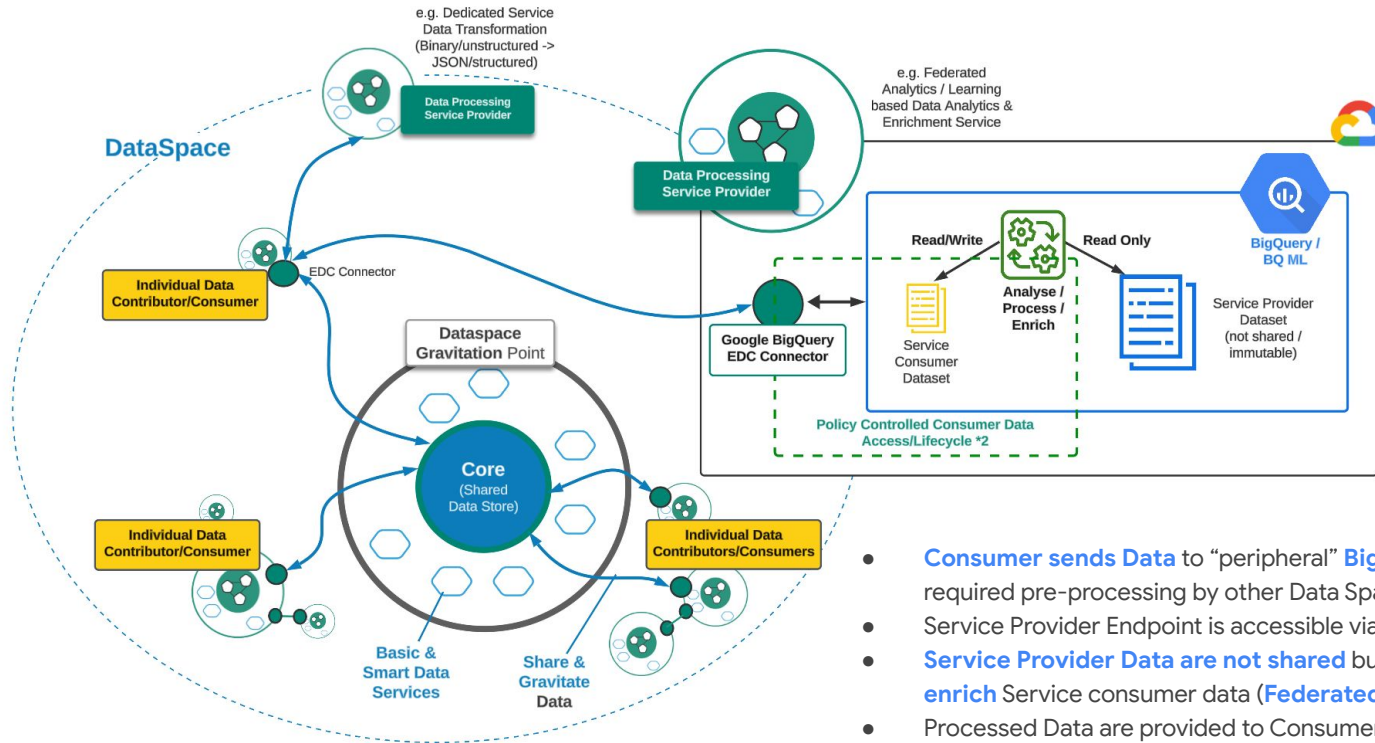
Gravitation vs. Expansion



- Holfelder, W., Mayer, A., Baumgart, T. (2022). Sovereign Cloud Technologies for Scalable Data Spaces. In: Otto, B., ten Hompe, M., Wrobel, S. (eds) Designing Data Spaces. Springer, Cham. https://doi.org/10.1007/978-3-030-93975-5_25

- **Enhance portability focused** Data Space Core Services with **interoperability focused “Peripheral Services”**
 - **Turn** general Data into (participant individual) Knowledge and Wisdom
 - **Combine** Data Space Data **with private / non-shareable** participant Data (e.g. Federated Learning)
 - **Contribute Data back** into the Core to **increase quality & value** of data



Complementary Data Space Example utilizing BigQuery*1



- **Consumer sends Data** to “peripheral” **BigQuery based Data Service** (might required pre-processing by other Data Space Service)
- Service Provider Endpoint is accessible via **Google BigQuery EDC Connector**
- **Service Provider Data are not shared** but can be used to help **analyze and enrich** Service consumer data (**Federated Analytics / Learning Approach**)
- Processed Data are provided to Consumer via Service Consumer Data Set
- **Value:** EDC compatible access and usage of service, no need to share data of service provider -> **highest level of Data Sovereignty**

*1 BigQuery Connector (BigQuery Extension) expected to be released Q2/24, subject to change, *2 Requirements and interoperability with existing dataspace implementation under review

Google Cloud offers choices to help protect data & workloads

	Connected
	Disconnected / Air-gapped

Organizations

Public Cloud



Does already support BigQuery

Commercial users & data

Core controls by default:
access, encryption, residency

Sovereign Controls

Highly sensitive, controlled and unclassified information (CUI)

Data Sovereignty controls
over encryption keys, admin access, service deployment location, partner services (logs reviews, BYOID)

Sovereign Trusted Partner Cloud

Highly sensitive data

Operational Sovereignty controls by trusted local partner over operations, while remaining "on-fabric"

Google Distributed Cloud

Classified, Secret, Top Secret data

Google Distributed Cloud Hosted (Google or partner operated)
Provides survivability and full segregation



Thank you.